DOCUMENT RESUME

ED 391 185 CS 509 116

AUTHOR Spicer, Karin-Leigh; Hanks, William E.
TITLE Multiple Measures of Critical Thinking Skills and Predisposition in Assessment of Critical Thinking.
PUB DATE Nov 95
NOTE 11p.; Paper presented at the Annual Meeting of the Speech Communication Association (81st, San Antonio, TX, November 18-21, 1995).
PUB TYPE Information Analyses (070) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Outcomes Assessment; Communication (Thought Transfer); *Critical Thinking; Higher Education; Logical Thinking; Problem Solving; Standardized Tests; *Student Evaluation; *Thinking Skills; Undergraduate Students
IDENTIFIERS Authentic Assessment; *Critical Thinking Skills Performance Assessment; *Multiple Measures Approach; Reflective Thinking

ABSTRACT

A panel of 46 experts from philosophy and education defines critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." At present, there are seven standardized critical thinking tests available, and several performance assessment approaches can be used as outcome measures within various subjects in communication. Standardized tests can provide useful information that is diagnostic and may help to guide instruction. However, multiple measures of critical thinking should be used in assessment. Critical thinking is not a general ability but rather a complex set of general and specific factors. Psychologists generally favor multiple measures of critical thinking because no single test covers the dimensions of a good conceptual definition of critical thinking. R. A. Ennis and S. P. Norris suggest that in lieu of appropriate multiple choice tests, open-ended assessment tests are needed; other measures could include interviews. College educators should first decide what students should be able to demonstrate and what they know and can do. Then, they should decide what to teach students. When educators are clear about the intended performance and results, they will have a set of criteria for selection of content. Then in devising their means of assessment, educators should consider guidelines concerning meaningful contexts in exams, novel situations, relevant products and performances, and the various levels of student ability. (Contains 21 references.) (TB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 391 185

Multiple Measures of Critical Thinking Skills and Predisposition in Assessment of Critical Thinking

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. L. Spicer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Karin-Leigh Spicer
William E. Hanks
Department of Communication
Wright State University
Dayton, Ohio 45435

CS509116

BEST COPY AVAILABLE

Multiple Measures of Critical Thinking Skills and Predisposition in Assessment of Critical Thinking

In 1990 Congress passed the Goals 2000: Educate America Act which included this goal: *The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially,* (US Department of Education, 1991). However, five years later Robert Ennis (1995), a major figure in the modern critical thinking movement, wrote: *Although critical thinking has often been urged as a goal of education throughout most of this century, not a great deal has been done about it* (p.179). Ennis (1990), among others (Halpern, 1993; Paul & Nosich, 1991; Facione, 1990) believes critical thinking assessment can be accomplished using a comprehensive definition of critical thinking, being clear about the purposes of assessment, and using multiple measures of critical thinking.

As Christ (1994) points out in Assessing Communication Education, assessment of programs should precede assessment of student outcomes (p.33). In other words faculty should have a clear idea of a given program's purpose. We assume that most communication programs intend to improve students' critical thinking skills. Further, we assume they want to define critical thinking broadly to include reflective judgement in thinking about ill-structured problems, discrete thinking skills such as the ability to spot fallacious reasoning, and strong predispositions for alternatives. Based on these assumptions, we offer a comprehensive definition of critical thinking, briefly describe several standardized critical thinking tests, argue that multiple measures of critical thinking should be used in assessment, and suggest several performance assessment approaches that can be used as outcome measures within various subjects in communication.

Although there are dozens of definitions of critical thinking, there is significant overlap in most of them (Halpern, 1993). The definition that captures the essence of most of those definitions, and, therefore, is the most comprehensive one is "Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction" (Facione, 1990). The definition is the product of a Delphi research project involving forty-six experts from philosophy and education, which says:

> *We understand critical thinking to be purposeful, self-regulatory judgement which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgement is based...The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider,*

clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. Thus, educating good critical thinkers means working toward this ideal.
It combines CT skills with those dispositions which consistently yield useful insights and which are the basis of a rational and democratic society (p.2).

This definition, though wordy and complex, includes both abilities and skills as well as important predispositions. Implicitly it also includes knowledge and assumes the more knowledgeable the student is potentially the better thinker s/he can become. Critical thinking, however, is difficult to measure because it covers so much territory.

Critical Thinking Tests

We know of seven tests of critical thinking that can be called general thinking tests (in the sense that they measure several kinds of thinking skills rather than only one or two). We have discovered only one instrument that measures critical thinking predispositions, such as the tendency to be analytical and to seek truth. We briefly describe each and note various strengths and weaknesses.

1. The Watson-Glaser Critical Thinking Appraisal (WGCTA)

Is the oldest and probably most widely used critical thinking test, and it has two parallel forms(that can be used in pre-posttest forms). It tests five types of skills: 1. inference; 2. recognition of assumptions; 3. deduction; 4. interpretation of data; and 5. evaluation of arguments (Watson & Glaser, 1980). The WGCTA has high reliability (.70 to .82) but some critics fault it for overreliance on deductive logic and for including inductive inference questions that are overly simplistic. As is the case with all general knowledge critical thinking tests, the content of questions may seem trivial.

2. The Cornell Conditional Reasoning Test, Form X

By Robert H. Ennis and Jason Miller (1985) contains seventy-two test items, and is intended for junior and senior high school and first year college students. Another form, **Level Z**, is intended for undergraduates, graduate students and adults. Form X tests the ability to tell whether a statement follows from the premises, something is reliable, an observation statement is reliable, a simple generalization is warranted, an hypothesis is specific, and whether a reason is relevant. Level X contains seventy-one multiple-choice items divided among four sections: 1. inductive inference; 2. credibility of sources and observation; 3. deduction; and 4. assumption identification. The Level Z test

contains fifty-two multiple-choice items with sections on deduction, meaning, credibility, induction, prediction, definition and unstated reasons, and assumptions. Reliability ratings are fairly good (Level X: .67 to .90; Level Z: .50 to .77), but the questions are sometimes too simplistic. As Ennis and Norris (1989) note, in administering this and other multiple-choice tests examinees should be allowed to explain why they answered as they did.

3. The Ross Test of Higher Cognitive Processes

This test is aimed at grades four through the college level. The test includes nine sections: 1. verbal analogies; 2. deduction, 3. assumption identification; 4. word relationships; 5. sentence sequencing; 6. interpreting answers to questions; 7. information sufficiency; 8. relevance in mathematics problems; and 9. analysis of attributes of complex stick figures (Ennis & Norris, 1990). Reliability estimates are exceptionally high (.92 for split-half and .94 for test-retest).

4. The New Jersey Test of Reasoning Skills

This test is intended for grades four through the college level. About one half of the test looks at classical syllogism and the meaning of categorical statements. The remainder of the test examines assumption identification, induction, good reasons, and distinguishing differences of kind and degree (Ennis & Norris, 1990). Reported reliability estimates range from .85 to .91, but these estimates are derived from non-college students. The test has been criticized for containing too many deductive logic questions and for answers keyed to specific background beliefs assumed to exist in the minds of examinees.

5. The Ennis-Weir Critical Thinking Essay Test

This test is geared for seventh through college level students. It is an essay format. The test includes getting the point, offering good reasons and assumptions, stating one's point, offering good reasons, seeing other possibilities, responding to\avoiding equivocation, irrelevance, circularity, reversal of an if-then relationship, overgeneralization, credibility problems, and the use of emotive language to persuade (Ennis & Norris, 1990). A scoring guide is provided, and interrater reliability estimates are .86 and .82.

6. The California Critical Thinking Skills Test

This test has two forms, A and B, that can be used in pre and posttest designs. It operationalizes the conceptual definition devised by the Delphi panel sponsored by the American Philosophical Association (APA, 1990). It is a thirty-four item, multiple choice test which targets those core critical thinking skills regraded to

be essential elements in a college education. The items range from those requiring an analysis of the meaning of a given sentence, to those requiring much more complex integration of CT skills. Reported reliability estimates are .70 for Form A and .71 for Form B. Some items are puzzle-like and about a third of test items contain deductive logic questions.

7. The California Critical Thinking Dispositions Inventory

The CCTDI, with seventy-five Likert style items, contain seven subscores and a total composite score (Facione & Facione, 1992). The seven scales are: 1. inquisitiveness; 2. open mindedness; 3. systematicity; 4. analyticity; 5. truth-seeking; 6. CT self-confidence; and 7. maturity. Examinees are asked to mark the degree to which they agree with statements such as, "If there were ten opinions on one side and one on the other, I'd go with the ten." Reported results of the inventory have consistently shown beginning college students are fairly strongly predisposed against truth-seeking. This instrument would seem quite useful in revealing attitudes of students toward critical thinking.

As we have noted, however each of the standardized instruments shares the basic weakness of reducing critical thinking to a set of responses. For that reason, among others, we strongly argue for the use of multiple measures of critical thinking, including, open-ended performance measures.

Multiple Measures

Critical thinking is not a general ability but a complex of general and specific factors (Follman, et al., 1969; Follman, et al., 1970). The APA Delphi Consensus definition of critical thinking seems clearly to recognize that point. And psychologists who have experimented with critical thinking, including Robert Sternberg (1987) favor multiple-measures of CT because no single test covers the dimensions of a good conceptual definition of critical thinking.

Ennis (1993) advises that for comprehensive assessment of critical thinking:

> *Unless appropriate multiple-choice tests are developed, open-ended assessment techniques are probably needed...In making your own, test, it is probably better that it be at least somewhat open-ended, anyway, since making good multiple-choice tests is difficult and time consuming...open-ended assessment is better adapted to do-it yourself makers and can be more comprehensive* (p. 184).

Ideally, according to Ennis and Norris (1990), other measures such as interviews can elicit useful information about how well students are thinking, e.g., question and answer sessions that allow for students to explain their thinking and for teachers to ask for follow up questions.

Performance assessments are good additions to pencil-paper tasks that call for a single correct answer. These types can include a variety of projects such as case-studies, research reports and portfolios. Performance-based assessment calls for demonstration of understanding and skill in applied, procedural, or open-ended settings (Baker, O'Neil, & Linn, 1993).

A good example of a performance measure, somewhat related to critical thinking performance, is described by Morreale (1994) in discussing the SCA's "The Competent Speaker" instrument for rating students' public speaking performance on eight speaking competencies. The student chooses and narrows a topic for a specific audience and occasion and must devise a thesis or purpose, provide supporting material, organize pronunciation, articulation and grammar, and physical movement (p.222). The instrument can be used to evaluate a speaker's performance in class, for placement in or out of classes, for instruction, and for assessment of programs or curricula.

Another argument for multiple-measures concerns the testing of predispositions separately. Facione et al. (1995) have noted every major theoretician since Dewey who wrote of critical thinking identified predisposition as important (pps. 1-25). Dewey argued:

> *If we were compelled to make a choice between these personal attributes and knowledge about the principles of logical reasoning together with some degree of technical skill in manipulating special logical processes, we should decide for the former* (1933).

Facione, et al. (1995) report a study of 587 new college students at a selective private university. Only thirteen percent showed positive predisposition toward all predispositions measured (i.e., truth-seeking, open-mindedness, analyticity, systematicity, CT-self-confidence, inquisitiveness, maturity). These were strong students academically. Yet, they almost all showed some opposition to truth-seeking. So far as we knew, this is the first study of predispositions in critical thinking. We need systematic evidence and profiles of students' attitudes in order to approach these attitudes directly. And conventional skills tests don't provide such evidence, for a student might display thinking skills only when required, yet be inclined not to do so at other times. Or a student might be strongly inclined to think well but simply lack the skill. We need to measure skill and tendency separately in order to teach thinking as both skill and habit of mind.

Guidelines for Creating Critical Thinking Assessments

For CT assessment tasks approach curriculum development backwards. First, decide what students should be able to demonstrate and what they know and can do. Second, decide what to teach students. This strategy can lead to coherence throughout the

entire curriculum. When educators have clarity about the intended performance and results they will have a set of criteria for selection of content, reducing aimless coverage and adjusting instruction en route and students will be able to grasp their priorities from the beginning. Educators should ask: *What is most essential for students to learn? Given what I want students to learn, what counts as evidence that they understand that?* These questions combine subject matter and critical thinking. But because available standardized CT tests are general and not subject-bound, such tests cannot provide evidence of how students think about critical thinking questions in meaningful context.

1. Meaningful Context

Good performance assessments are more contextualized than traditional tests. A question we should ask is: How will students use CT skills in the larger world? The American Educational Research Association, American Psychological Association and The National Council on Measurement in Education (1985) specified the following criteria for performance assessments. They should: 1. Have meaning for students and teachers and motivate high performance; 2. Require the demonstration of complex cognition (e.g. problem solving, knowledge representation, explanation); 3. Exemplify current standards of content or subject matter quality; 4. Minimize the effects of ancillary skills that are irrelevant to the focus of the assessment; and 5. Possess explicit standards for rating or judgment. The standards that apply generally to performance assessment should apply to critical thinking assessment. How? Specific examples are needed that are highly structured with clear criteria, such as the performance assessment described by Morreale. In interpersonal communication situations, for example, case-studies could be used and student performances rated on specific criteria. In mass communication, to cite another example, students might be asked to present a synthesis of media effects and to argue for or against a specific government policy on media regulations.

Performance measures of thinking can be used as students progress through a given program, for example, during the junior and senior years. Unlike standardized non-subject specific critical thinking tests, however, performance measures should not be used as pretest measures as students enter the major. The reason should be obvious--students cannot be assumed to possess the required knowledge to handle context-bound thinking tasks before they have taken the courses that provide subject-specific critical thinking.

2. Thinking Process in Performance Assessments

Ask students to actually use knowledge, to thoughtfully address situations that are novel to them. This is not to say that certain discrete skills, such as the ability to identify assumptions, cannot be taught as discrete skills. It is merely to say that the value in thinking skills is whether they transfer to meaningful and unique contexts. However, a skill that is cut free

from content and context is measurable and teachable, but of only limited value. Generic CT skills and their assessments do not reveal the depth and breadth of student knowledge. For example, Paul and Nosich (1991) list seventeen critical thinking abilities. These include refining generalizations and avoiding oversimplification, clarifying issues, conclusions and beliefs, generating or assessing solutions, and analyzing or evaluating actions or policies. Such skills can be learned, but they probably are best learned if placed in meaningful context.

3. Appropriate Product or Performance

Avoid using products or performances that don't relate to the content of what is being assessed. Sometimes students as well as educators can get caught up in the product and lose sight of what they're actually intending to show with the product.

4. How Material is Taught

Good CT assessments are designed to guide, not limit, instruction. They should not infringe on educators' abilities to choose particular methods and to design lessons and courses in ways that reflect the best available research and which are best suited to their students' needs.

5. Multiple Performance Levels

It is not realistic to expect all students to meet the same standard. Multiple standards can set expectations to match different aspiration and achievements. A single standard would either have to be set low enough for most students to pass or too high for many to reach. Setting standards that are within reach but still require hard work can stretch students to their potential. For example, require all students to meet a common standard for obtaining their degree, but also create a higher standard for students who attain that initial level earlier or who wish to qualify for more selective higher education.

Conclusion

We have argued that multiple-measures are needed to assess students' critical thinking abilities. Standardized critical thinking tests can provide useful information that is diagnostic and may help to guide instruction. But instruction cannot be limited to teaching the skills measured by the instruments. To measure thinking skills that require application of knowledge requires specially designed tasks, including performance tasks for which there are specified outcome criteria but for which there can be established general evaluation rubrics. Finally, we should emphasize that the purpose of assessment should be to improve instruction, learning and programs, and all data in that context should regard both as formative and summative.

Endnotes

American Philosophical Association (1990). *The California Critical Thinking Skills Test.*

Baker, E.L., O'Neil, H.F., & Linn, R.L. (1993). *Policy and validity prospect performance-based assessment.* American Psychologist, 48(12), 1210-1218.

Christ, W.G. (1994). *Mission statements, outcomes, and the new Liberal Arts.* In Christ, W.G. (Ed.), Assessing Communication Education: A Handbook for Media, Speech, & Theater Educators, (pg. 479). Norwood, N.J.: Lawrence Earlbaum Associates.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educational process.* Lexington, MA: Heath.

Ennis, R.H. (1993). *Critical thinking assessment.* Theory into Practice. 32(3), 179, 186.

_____, & Norris, S.P. (1990). *Critical thinking assessment: Status, issues, needs.* In Algina, J. & Leggs, S. (Eds.), Cognitive assessment of language and math outcomes (pp. 1-42). Norwood, N.J.: Ablex Publishing.

_____, & Millman, J. (1985). *Cornell Critical Thinking Test, Level X.* Pacific Grove, CA: Midwest Publications.

_____(1985). *Cornell Critical Thinking Test, Level Z.* Pacific Grove, Ca: Midwest Publications.

Facione, P.A. (1995). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* ERIC-ED 315 423.

_____, Sanchez, C.A., Facione, N.C., Gainen, J. (1995). *The disposition toward critical thinking.* The Journal of General Education, 44(1), 1-24.

_____, & Facione, N.C. (1992). *The California Critical Thinking Dispositions Inventory and Test Manual.* Millbrae, CA: California Academic Press.

_____, P.A. (1991). *Using the California Critical Thinking Skills Test in research, evaluation, and assessment.* Millbrae, CA: The California Press.

Follman, J., Brown, L., & Burg, E. (1970). *Factor analysis of critical thinking, logical reasoning, and English subtests.* The Journal of Experimental Education, 34(4), 11-16.

_____, Hernandex, D., & Miller, W. (1969) *Canonical correlation of scholastic aptitude, and critical thinking subtests.* The Journal of Experimental Education, 38(1), 49-53.

Halpern, D.F. (1993). *Assessing the effectiveness of critical thinking instruction.* The Journal of General Education, 42(4), 238-254.

Morreale, S. (1994) *Public Speaking.* In Christ, W.G. (Ed.) Assessing Communication Education: A Handbook for Media, Speech, & Theater Educators, (pp. 219-232). Norwood, N.J.: Lawrence Earlbaum Associates.

Paul, R.W., & Nosich, G.M. (1991). *A proposal for the national assessment of higher-order thinking.* Washington D.C.: The US Department of Education Office of Educational Research and Improvement National Center for Education Statistics.

Sternberg, R.J. (1987). *Questions and answers about the nature and teaching of thinking skills.* In Baron, J.B., Sternburg, R.J. (Eds.), Teaching Thinking Skills: Theory and Practice (pp. 250-262). N.Y.: W.H. Freeman.

Watson, G., & Glaser, E. (1980). *Watson-Glaser Critical Thinking Appraisal.* San Antonio, TX: Psychological Corporation.

U.S. Department of Education (1991). American 2000: An Education Strategy. Washington, D.C.: American Educational Research Association

U.S. Department of Education (1990). *National goals for Education.* Washington D.C.: Government Printing Office.